DOI:10.11931/guihaia.gxzw201905048

基于 5 993 个核基因的被子植物系统发育关系研究

金鑫[1,2]，程书[1,2]，杨拓[3]，余慷[1,2]，段肖霞[1,4]，倪雪梅[1,4]，李世明[1,2]，张耕耘[1,4*]

（1. 深圳华大生命科学研究院，广东 深圳 518083；2. 深圳市华大农业应用研究院，广东 深圳 518120；3. 国家基因库，广东 深圳 518120；4. 基因组学农业部重点实验室，广东 深圳 518083）

摘要：系统发育关系的构建对被子植物分类及进化研究非常重要。长期以来，被子植物系统发育的研究，大多使用质体基因、线粒体基因或少数保守的单拷贝核基因。本研究从已注释基因组或转录组中搜集 88 种被子植物（包含 58 目）的核基因集；通过对其进行同源基因聚类及去旁系同源基因，获得了 5 993 个一对一的直系同源基因家族（即对于每个基因家族，每种植物最多一条序列，最少包含 50 个物种）；使用截取各种不同数目基因集的 DNA 或氨基酸序列，采用串联法（concatenation）和溯祖法（coalescence），共构建了 20 棵进化树。比较这些进化树，虽然大部分结果支持 APG IV 中描述的被子植物主要支系之间的关系（(真双子叶植物，单子叶植物)，木兰类植物），但真双子叶植物内部各目分支的演化关系与 APG IV 有一个很大的不同，即本研究认为檀香目和石竹目是蔷薇类植物的姊妹群。基于这些进化树，估算了被子植物各目分支的分化时间，结果表明被子植物的起源时间为 237.78 百万年前（95%置信区间为 202.6~278.08），与主流观点认为的 225~240 百万年前一致。本研究为构建进化树提供了一种可行性策略，这种方法允许使用基因数目更多而计算速度更快。

关键词：系统发育关系，被子植物，核基因，同源基因聚类，串联法，溯祖法，分化时间

中图分类号：Q949.4　　　　**文献标识码：**A

Reconstruction of angiosperm phylogeny based on 5 993 nuclear genes

JIN Xin[1,2], CHENG Shu[1,2], YANG Tuo[3], YU Kang[1,2], DUAN Xiaoxia[1,4], NI Xuemei[1,4], LI Shiming[1,2], ZHANG Gengyun[1,4*]

（1. *BGI-Shenzhen*, Shenzhen 518083, Guangdong, China; 2. *BGI Institute of Applied Agriculture*, Shenzhen 518120, Guangdong, China; 3. *China National Gene Bank*, Shenzhen 518120, Guangdong, China; 4. *Key*

基金项目：国家科技支撑计划（2015BAD02B01-7）；广东省农作物核心资源开发应用企业重点实验室（2011A091000047）；深圳市科技计划项目（JCYJ20150831201123287）；深圳作物分子设计聚合育种工程实验室提升项目（深发改[2015]946 号） [Supported by National Science and Technology Support Program (2015BAD02B01-7); Key Laboratory of Crop Core Resources Development and Application Enterprises of Guangdong (2011A091000047); Science and Technology Program of Shenzhen (JCYJ20150831201123287); Molecular Design and Polymerization Breeding Engineering Laboratory of Shenzhen (Shenfagai[2015]946)]。

作者简介：金鑫（1987-），男，湖北潜江人，硕士研究生，研究方向为作物遗传育种与生物信息学，（E-mail）jinxin1@genomics.cn。

[*] **通信作者：**张耕耘，博士，研究员，研究方向为基因组学与植物育种，（E-mail）zhanggengyun@genomics.cn。

Laboratory of Genomics, Ministry of Agriculture, BGI-Shenzhen, Shenzhen 518083, Guangdong, China）

Abstract: Construction of phylogeny is important to classification and research of angiosperms. For a long time, angiosperm phylogeny has been analysed using plastid genes, mitochondrial genes or a few conserved single-copy nuclear genes. Here, we collected nuclear gene sets of 88 species of angiosperm (contains 58 orders) from annotated genomes or transcriptomes. By using a combined homology- and phylogeny tree-based approach, we obtained a total of 5 993 one-to-one ortholog groups (one sequence of each species for each ortholog group), each of which was represented by at least 50 species. Then, a total of 20 species trees were reconstructed using different combination of reconstruction methods (concatenation-based and coalescence-based) and sequence type (nucleotide or amino acid) for gene data sets with different gene occupancy values. Most of the resulting topologies support the relationships of the major clades of angiosperm as described in APG IV, but present different deep relationships among major clades in eudicots phylogeny such as the placement of Santalales and Caryophyllales as sisters to Rosids. We estimated the divergence times of the major clades of angiosperm and concluded that the origin of angiosperm is about 237.78 million years ago(95% confidence interval is 202.6~278.08), which is in accordance with the previously accepted 225~240 million years ago. This study provided an efficient strategy for building phylogenetic trees using thousands of genes with ultrafast calculation.
Key words: phylogeny, angiosperms, nuclear genes, ortholog inference, concatenation, coalescence, divergence time

系统发育树的正确构建对植物分类及进化研究非常重要。进化树构建的准确度主要受以下因素的影响。其一，所使用的数据集的种类及大小。不仅使用形态性状数据、质体基因、线粒体基因及核基因序列建立的进化树不一样（Endress & Doyle, 2009; Ruhfel et al., 2014; Soltis et al., 2011; Zeng et al., 2014），使用全长核酸序列、或仅使用基因密码子某个位点的核酸序列及氨基酸序列所构建的进化树也有所不同（Wickett et al., 2014）。其二，构建树的方法及模型。方法有串联法（concatenation）和溯祖法（coalescence）：串联法是将所有基因串联作为一个整体，使用软件 RAxML（Stamatakis, 2014）或 iqtree（Nguyen et al., 2015）构建系统发育树；溯祖法是先对每个基因建树，再使用软件 ASTRAL（Zhang et al., 2017）建立所有基因树的共有树（Wickett et al., 2014）。而构建系统发育树使用的模型更是多种多样，如核酸模型 GTR、HKY、JC、F81、K2P、K3P、K81uf 等，蛋白质模型 LG、Poisson、cpREV、mtREV、Dayhoff、mtMAM、JTT、WAG 等（Nguyen et al., 2015）。

被子植物是植物界最高等且种类最多的一类，它们在地球上占据着绝对优势。现在已报道被子植物有 352 000 种（http://www.theplantlist.org/），属于 416 科和 64 目，各目之间的演化关系一直是研究的热点和争论的焦点。被子植物除了最基部的三个目：无油樟目（Amborellales）、睡莲目（Amborellales）和木兰藤目（Austrobaileyales），又称 ANITA 组，其余的（99.95%）可以分为五类：木兰类植物（magnoliids）、单子叶植物（monocots）、真双子叶植物（eudicots）、金粟兰科（Chloranthaceae）和金鱼藻科（Ceratophyllaceae）。这五类的系统演化拓扑关系一直存在争论，Zeng et al.(2014) 总结了已经发表的五种主要的拓扑关系（图 1:A-E），其中 A 是最主流的，也是 APG IV（THE ANGIOSPERM PHYLOGENY GROUP, 2016）的拓扑结构。Soltie et al.(2011) 使用 17 个基因串联（包括质体基因、线粒体基因和核基因）为 640 种植物构建的系统发育进化树，和 Ruhfel et al.(2014) 使用 78 个质体基因串联为 360 种植物构建的进化树，支持主流 A 拓扑结构。Wickett et al.(2014) 使用 674

个核基因串联为 92 种植物构建的进化树，和 Zeng et al.(2014)使用 59 个核基因串联为 61 种植物构建的进化树，支持 B 拓扑结构。除此之外，Qiu et al.(2010)使用 4 个线粒体基因为 380 种植物构建的进化树，支持 C 拓扑结构；Endress & Doyle(2009)使用形态性状构建的进化树，支持 D 拓扑结构；Zhang et al.(2012)使用 5 个核基因为 91 种植物构建的进化树，支持 E 拓扑结构。

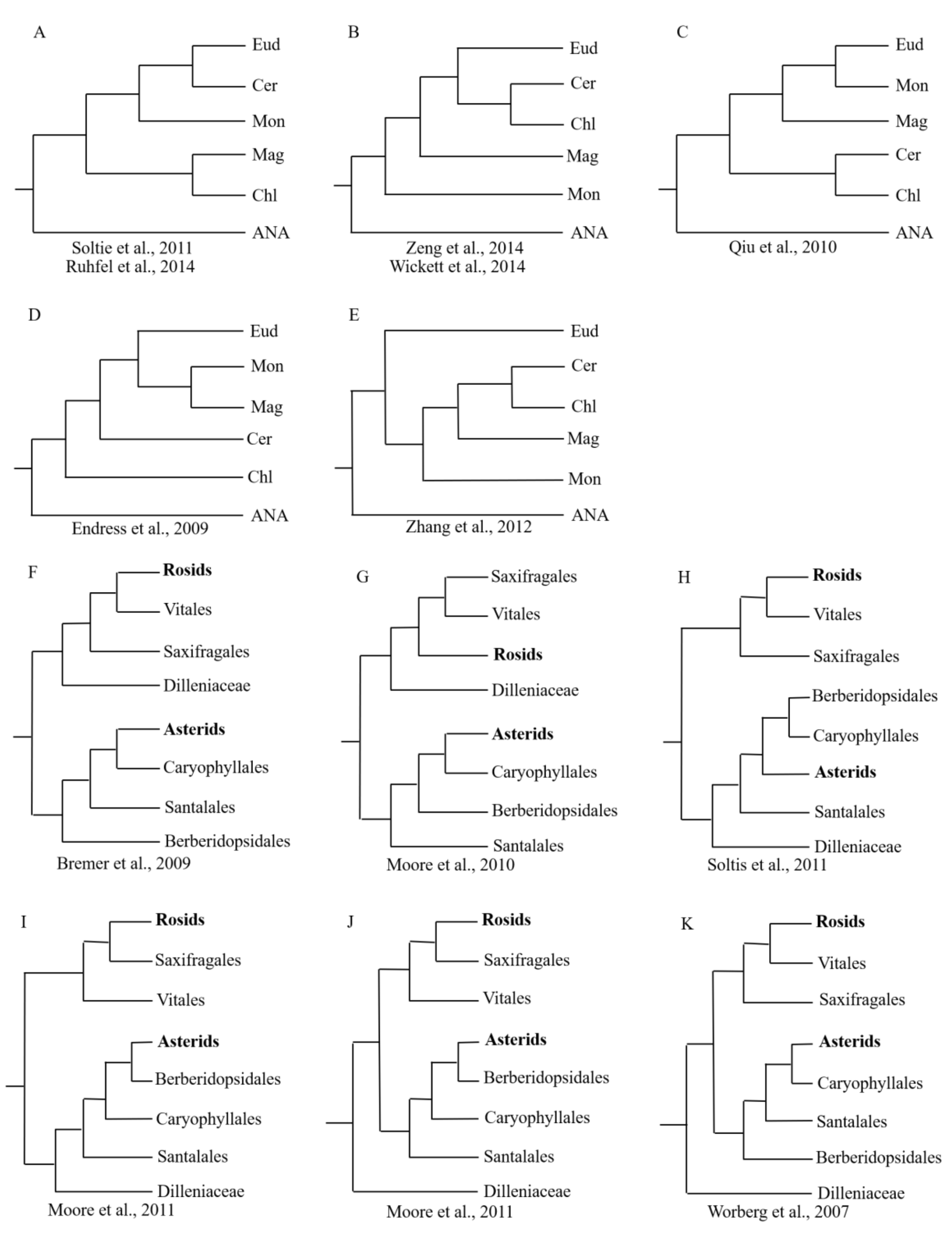
去掉金粟兰科和金鱼藻科后，单子叶植物、木兰类植物、真双子叶植物之间的系统发育关系有三种：((真双子叶植物，单子叶植物)，木兰类植物)；((真双子叶植物，木兰类植物)，单子叶植物)；((单子叶植物，木兰类植物)，真双子叶植物)。Lu et al.(2018)使用 4 个质体基因和 1 个线粒体基因分析了 5 864 种中国被子植物（几乎包括所有中国地区被子植物）的系统发育关系，其构建的进化树支持拓扑结构((真双子叶植物，单子叶植物)，木兰类植物)。Chen et al.(2019)发布了木兰类植物鹅掌楸（*Liriodendron*）基因组，使用其 502 个核基因及溯祖法为 18 种植物构建的进化树，同样支持拓扑结构((真双子叶植物，单子叶植物)，木兰类植物)。Chaw et al.(2019)发布了另一个木兰类植物牛樟（stout camphor tree）基因组，使用其 211 个核基因为 13 种植物构建的进化树，支持拓扑结构((真双子叶植物，木兰类植物)，单子叶植物)。Li et al.(2019)使用 2881 种被子植物的质体基因组的 80 个基因重建了被子植物高分辨率的系统发育树，支持拓扑结构((真双子叶植物，单子叶植物)，木兰类植物)。从上述已有的研究中，我们发现，使用核基因串联法建立的进化树基本都支持拓扑结构((真双子叶植物，木兰类植物)，单子叶植物)，使用核基因溯组法、质体和线粒体基因建立的进化树基本都支持拓扑结构((真双子叶植物，单子叶植物)，木兰类植物)。

真双子叶植物内部各目的系统发育关系也存在争论（图 1:F-K），真双子叶植物除了最基部的毛茛目（Ranunculales）、山龙眼目（Proteales）、昆栏树目（Trochodendrales）、黄杨目（Buxales）和洋二仙草目（Gunnerales），其余的可以分为两类：蔷薇类植物（Rosids）和菊类植物（Asterids）。这两类植物的基部有 6 个目的系统发育关系比较混乱，即五桠果目（Dilleniales）、虎耳草目（Saxifragales）、葡萄目（Vitales）、檀香目（Santalales）、智利藤目（Berberidopsidales）及石竹目（Caryophyllales）。Zeng et al.(2017)总结了已经发表的六种主要的拓扑关系（图 1:F-K），其中 K 是 APG IV 中所认可的拓扑结构。Moore et al.(2010)使用 83 个质体基因为 86 种植物构建的进化树，支持"五桠果目是蔷薇类植物的姊妹群"；Soltie et al.(2011)等使用 17 个基因串联（包括质体基因、线粒体基因和核基因）为 640 种植物构建的进化树，和 Moore et al.(2011)使用质体 IR 序列为 87 种植物构建的进化树，支持"五桠果目是菊类植物的姊妹群"；Worberg et al.(2007)等使用五个基因组区域序列为 56 种植物构建的进化树，和 Moore et al.(2011)使用质体 IR 序列为 244 种植物构建的进化树，及 APG IV 都支持"五桠果目是蔷薇类植物和菊类植物共同的姊妹群"。大部分研究都支持"葡萄目和虎耳草目是蔷薇类植物的姊妹群，智利藤目、檀香目和石竹目是菊类植物的姊妹群"（Moore et al., 2011, 2010; Worberg et al., 2007; Yang et al., 2015）；Zeng et al.(2017)使用 504 个核基因并联为 100 种植物构建的进化树，支持"檀香目和智利藤目是蔷薇类植物的姊妹群"。

被子植物的起源及进化一直是植物学界研究和争论的热点。在古生物学界，很长时期内，被子植物的最早化石记录都是白垩纪 125 百万年前，也是最早的真双子叶植物化石记录（Herendeen, 1995）。Fu et al.(2018)发现了早侏罗纪地层（约 175 百万年前）中的"南京花"，其具有花萼、花瓣、雌蕊，有明显的杯托、下位子房上位花、树状的花柱，其种子/胚珠确实是被完全包裹着，子房壁将种子与外界完全隔绝，这都满足了被子植物判断标准。"南京花"的发现，将被子植物最早化石记录向前推进了约 5 000 万年，并填补了被子植物化石记录（125 百万年前）与分子钟推算时间（225~240 百万年前）之间的"侏罗纪空缺"（Jurassic gap）（Li et al., 2019）。目前，大多数基于系统进化树的被子植物分化时间估计研究，都认为被子植物的起源为三叠纪 225~240 百万年前(Magallon, 2010; Mandel, 2019; Smith et al., 2010;

Zeng et al., 2014），这与起传粉作用的核心植食性鳞翅目昆虫的起源时间（约 230 百万年前）一致（Li et al., 2019; Zeng et al., 2014）。

本研究使用超过 5 000 个核基因的核酸及蛋白序列，用两种进化树构建方法分析了 88 种被子植物的系统发育关系（包括 87 科 58 目），并对各进化分支的分化时间进行了估计（总流程如图 2）。为了得到准确可靠的被子植物系统发育进化树，我们对 5 000 多个核基因进行了拆分，得到了包含不同基因数量的多个数据集，并对各个数据集进行系统发育树的构建，最后比较了所得到的 20 棵系统发育进化树之间的一致性。



注：**A-E.** 五类被子植物间（金粟兰科(Chl)、金鱼藻科(Cer)、木兰类植物(Mag)、单子叶植物(Mon)及真双子叶植物(Eud)）5 种代表性拓扑结构；**F-K.** 真双子叶植物内部各目间 6 种代表性拓扑结构。

Note：**A-E**. Five representative topologies among eudicots (Eud), monocots (Mon), magnoliids (Mag), Ceratophyllaceae (Cer) and Chloranthaceae (Chl); **F-K**. Six representative topologies among eudicots.

图 1 不同拓扑结构的被子植物演化关系

Fig.1 Various topologies of angiosperm phylogeny

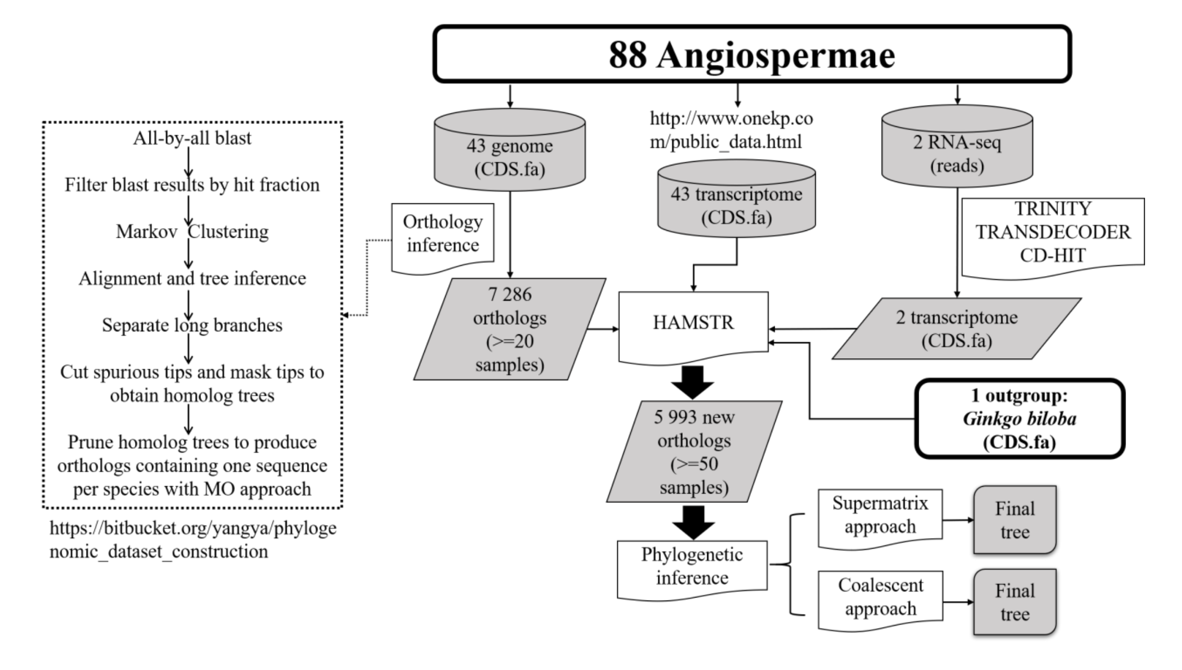


图 2 被子植物系统演化关系重建总流程

Fig.2 The overall workflow of angiosperm phylogeny reconstruction

1 材料和方法

1.1 材料

我们收集了 1 个裸子植物（*Ginkgo biloba* 作为外类群）基因组、43 个被子植物基因组（主要来自 NCBI 和 PHYTOZOME 数据库）、43 个被子植物已拼接转录组（http://www.onekp.com/public_data.html ）及 2 个被子植物 RNA-seq 数据（其中无叶莲 *Petrosavia sakurai* 是本研究测序的物种），其中被子植物共包含 87 科 58 目（表 1）。

表 1 本研究所用的 89 个植物物种及数据来源

Table 1 The list of the 89 plants used in this study and the data source

物种 Species	数据来源类型 Data origin type	缩写 Abbreviation	目 Order	数据来源 Data origin
银杏 *Ginkgo biloba*	Genome	GGIBI	裸子植物门 Gymnosperm	http://gigadb.org/dataset/100613
猕猴桃 *Actinidia chinensis*	Genome	GACCH	杜鹃花目 Ericales	ncbi
无油樟 *Amborella trichopoda*	Genome	GAMTR	无油樟目 Amborellales	phytozome
菠萝 *Ananas comosus*	Genome	GANCO	禾本目 Poales	ncbi
深圳拟兰 *Apostasia shenzhenica*	Genome	GAPSH	天门冬目 Asparagales	ncbi
拟南芥 *Arabidopsis thaliana*	Genome	GARTH	十字花目 Brassicales	tair
芦笋 *Asparagus officinalis*	Genome	GASOF	天门冬目 Asparagales	phytozome
甜菜 *Beta vulgaris*	Genome	GBEVU	石竹目 Caryophyllales	ncbi

胡桃 *Juglans regia*	Genome	GJURE	壳斗目 Fagales	ncbi
喜树 *Camptotheca acuminata*	Genome	GCAAC	山茱萸目 Cornales	https://datadryad.org/resource/doi:10.5061/dryad.nc8qr
番木瓜 *Carica papaya*	Genome	GCAPA	十字花目 Brassicales	phytozome
长春花 *Catharanthus roseus*	Genome	GCARO	龙胆目 Gentianales	https://datadryad.org/resource/doi:10.5061/dryad.hs593
土瓶草 *Cephalotus follicularis*	Genome	GCEFO	酢浆草目 Oxalidales	ncbi
甜橙 *Citrus sinensis*	Genome	GCISI	无患子目 Sapindales	phytozome
中粒咖啡 *Coffea canephora*	Genome	GCOCA	龙胆目 Gentianales	http://coffee-genome.org/download
胡萝卜 *Daucus carota*	Genome	GDACA	伞形目 Apiales	phytozome
龙眼 *Dimocarpus longan*	Genome	GDILO	无患子目 Sapindales	http://gigadb.org/dataset/view/id/100276
几内亚薯蓣 *Dioscorea rotundata*	Genome	GDIRO	薯蓣目 Dioscoreales	http://genome-e.ibrc.or.jp/home/bioinformatics-team/yam
苦荞麦 *Fagopyrum tataricum*	Genome	GFATA	石竹目 Caryophyllales	http://www.mbkbase.org/Pinku1
小果野蕉 *Musa acuminata*	Genome	GMUAC	姜目 Zingiberales	ncbi
画眉草 *Eragrostis tef*	Genome	GERTE	禾本目 Poales	https://genomevolution.org/CoGe/GenomeInfo.pl?gid=22790
巨桉 *Eucalyptus grandis*	Genome	GEUGR	桃金娘目 Myrtales	phytozome
欧洲白蜡树 *Fraxinus excelsior*	Genome	GFREX	唇形目 Lamiales	http://www.ashgenome.org/dup_annot11
大豆 *Glycine max*	Genome	GGLMA	豆目 Fabales	phytozome
向日葵 *Helianthus annuus*	Genome	GHEAN	菊目 Asterales	phytozome
牵牛花 *Ipomoea nil*	Genome	GIPNI	茄目 Solanales	ncbi
伽蓝 *Kalanchoe fedtschenkoi*	Genome	GKAFE	虎耳草目 Saxifragales	phytozome
博落回 *Macleaya cordata*	Genome	GMACO	毛茛目 Ranunculales	ncbi
苦瓜 *Momordica charantia*	Genome	GMOCH	葫芦目 Cucurbitales	ncbi
中国莲 *Nelumbo nucifera*	Genome	GNENU	山龙眼目 Proteales	ncbi
海枣 *Phoenix dactylifera*	Genome	GPHDA	棕榈目 Arecales	ncbi
毛果杨 *Populus trichocarpa*	Genome	GPOTR	金虎尾目 Malpighiales	phytozome
委陵菜 *Potentilla micrantha*	Genome	GPOMI	蔷薇目 Rosales	http://gigadb.org/dataset/100407
黄花九轮草 *Primula veris*	Genome	GPRVE	杜鹃花目 Ericales	https://www.datadryad.org/resource/doi:10.5061/dryad.2s200
石榴 *Punica granatum*	Genome	GPUGR	桃金娘目 Myrtales	ncbi
栎栎 *Quercus lobata*	Genome	GQULO	壳斗目 Fagales	https://valleyoak.ucla.edu/genomicresources/

橡胶树 *Hevea brasiliensis*	Genome	GHEBR	金虎尾目 Malpighiales	ncbi
丹参 *Salvia miltiorrhiza*	Genome	GSAMI	唇形目 Lamiales	http://gigadb.org/dataset/100164
马铃薯 *Solanum tuberosum*	Genome	GSOTU	茄目 Solanales	phytozome
紫萍 *Spirodela polyrrhiza*	Genome	GSPPO	泽泻目 Alismatales	phytozome
可可 *Theobroma cacao*	Genome	GTHCO	锦葵目 Malvales	phytozome
葡萄 *Vitis vinifera*	Genome	GVIVI	葡萄目 Vitales	phytozome
大枣 *Ziziphus jujuba*	Genome	GZIJU	蔷薇目 Rosales	ncbi
大叶藻 *Zostera marina*	Genome	GZOMA	泽泻目 Alismatales	phytozome
美洲菖蒲 *Acorus americanus*	Transcriptome	MTII	菖蒲目 Acorales	onekp
青荚叶 *Helwingia japonica*	Transcriptome	QACK	冬青目 Aquifoliales	onekp
木兰藤 *Austrobaileya scandens*	Transcriptome	FZJL	木兰藤目 Austrobaileyales	onekp
红花八角 *Illicium floridanum*	Transcriptome	VZCI	木兰藤目 Austrobaileyales	onekp
鳞枝树 *Aextoxicon punctatum*	Transcriptome	QUTB	智利藤目 Berberidopsidales	onekp
红金藤 *Berberidopsis beckleri*	Transcriptome	HAEU	智利藤目 Berberidopsidales	onekp
圆锥梅藤 *Mertensia paniculata*	Transcriptome	DKFZ	紫草目 Boraginales	onekp
加洲蓝钟 *Phacelia campanularia*	Transcriptome	YQIJ	紫草目 Boraginales	onekp
锦熟黄杨 *Buxus sempervirens*	Transcriptome	IWMW	黄杨目 Buxales	onekp
白桂皮 *Canella winterana*	Transcriptome	DDEV	白桂皮目 Canellales	onekp
林仙 *Drimys winteri*	Transcriptome	WKSU	白桂皮目 Canellales	onekp
玉女樱 *Crossopetalum rhacoma*	Transcriptome	IHCQ	卫矛目 Celastrales	onekp
金鱼藻 *Ceratophyllum demersum*	Transcriptome	NPND	金鱼藻目 Ceratophyllales	onekp
红茎蛔囊花 *Ascarina rubricaulis*	Transcriptome	WZFE	金粟兰目 Chloranthales	onekp
水苏 *Stachyurus praecox*	Transcriptome	VYGG	燧体木目 Crossosomatales	onekp
美国省沽油 *Staphylea trifolia*	Transcriptome	PTLU	燧体木目 Crossosomatales	onekp
五桠果 *Dillenia indica*	Transcriptome	EHNF	五桠果目 Dilleniales	onekp
日本珊瑚树 *Viburnum odoratissimum*	Transcriptome	HLJG	川续断目 Dipsacales	onekp
金银花 *Lonicera japonica*	Transcriptome	GSZA	川续断目 Dipsacales	onekp
红叶风眼莲 *Escallonia rubra*	Transcriptome	CLMX	南鼠刺目 Escalloniales	onekp
杜仲 *Eucommia ulmoides*	Transcriptome	SZUO	绞木目 Garryales	onekp
花茎草 *Francoa appendiculata*	Transcriptome	HDWF	牻牛儿苗目 Geraniales	onekp
斑点老鹳草 *Geranium maculatum*	Transcriptome	YGCX	牻牛儿苗目 Geraniales	onekp
长萼大叶草 *Gunnera manicata*	Transcriptome	XMQO	洋二仙草目 Gunnerales	onekp
瘿椒树 *Tapiscia sinensis*	Transcriptome	WWKL	十齿花目 Huerteales	onekp
锦葵叶刺核藤 *Pyrenacantha malvifolia*	Transcriptome	QZZU	茶茱萸目 Icacinales	onekp
钩药茶 *Oncotheca balansae*	Transcriptome	PVGM	茶茱萸目 Icacinales	onekp
美国蜡梅 *Calycanthus floridus*	Transcriptome	FALI	樟目 Laurales	onekp
美檫树 *Sassafras albidum*	Transcriptome	ABSS	樟目 Laurales	onekp
阿福花状旱叶草 *Xerophyllum asphodeloides*	Transcriptome	AFLV	百合目 Liliales	onekp
锯齿绿荆棘 *Smilax bona-nox*	Transcriptome	MWYQ	百合目 Liliales	onekp
刺果番荔枝 *Annona muricata*	Transcriptome	YZRI	木兰目 Magnoliales	onekp
肉豆蔻 *Myristica fragrans*	Transcriptome	OBPL	木兰目 Magnoliales	onekp
黄瑞香 *Daphne giraldii*	Transcriptome	PUDI	锦葵目 Malvales	onekp

圆叶肋果萍蓬草 *Nuphar advena*	Transcriptome	WTKZ	睡莲目 Nymphaeales	onekp
露兜树 *Xerophyta villosa*	Transcriptome	QOXT	露兜树目 Pandanales	onekp
马蹄香 *Saruma henryi*	Transcriptome	QDVW	胡椒目 Piperales	onekp
墨西哥胡椒 *Piper auritum*	Transcriptome	MUNP	胡椒目 Piperales	onekp
银桦 *Grevillea robusta*	Transcriptome	GRRW	山龙眼目 Proteales	onekp
框东坚果 *Santalum acuminatum*	Transcriptome	RSPO	檀香目 Santalales	onekp
昆栏树 *Trochodendron aralioides*	Transcriptome	SWOH	昆栏树目 Trochodendrales	onekp
蔓茎刺球果 *Krameria lanceolata*	Transcriptome	ZHMB	蒺藜目 Zygophyllales	onekp
蒺藜 *Tribulus eichlerianus*	Transcriptome	KVAY	蒺藜目 Zygophyllales	onekp
露水草 *Cyanotis arachnoidea*	RNA-SEQ	TCYTR	鸭跖草目 Commelinales	https://www.ncbi.nlm.nih.gov/sra/SRP144398
无叶莲 *Petrosavia sakurai*	RNA-SEQ	TPETR	无叶莲科 Petrosaviaceae	This study

1.2 基于基因组序列的直系同源基因鉴定

我们使用 Yang & Smith(2014)报道的方法，对 43 个植物基因组的基因集进行同源基因聚类分析。先使用软件 BLASTN v2.6.0+ 对 43 个基因集 CDS 序列进行 all-by-all blast，每条序列取最佳的 1 000 条比对结果，去掉比对长度小于 1/3 总长的序列，修剪未比对上的末端序列。再使用 MCL 软件（Van, 2000）进行同源基因聚类（inflation value = 1.4），去除少于 20 个植物的基因家族，剩余基因家族使用 MAFFT v7.310 软件（Katoh & Standley, 2013）进行多序列比对（maximum iterative refinement cycles = 1 000），使用 PHYUTILITY v2.2.6 软件（Smith & Dunn, 2008）修剪缺失率大于 90%的位点，使用软件 RAXML v8.2.11（Stamatakis, 2014）对修剪后的多序列比对数据估算系统进化树（model = GTRCAT）。最后修剪掉进化树上的所有旁系同源基因枝，修剪枝长大于 0. 6 的枝、比姐妹枝长十倍的末端枝，单源且全部同样品的枝只保留一个，修剪枝长比预期碱基替换率大 0. 3 倍的内部枝，再使用 MO 方法（Yang & Smith, 2014）去除所有剩余的旁系同源枝，获得 one-to-one 同源基因家族（即每个样品最多一条序列），只保留大于 20 个样品的基因家族。

1.3 转录组及外类群数据处理

我们对两个来自两个科（无叶莲科 *Petrosavia sakurai* 和鸦跖草科 *Cyanotis arachnoidea*）的 RNA-seq 数据从头拼接。首先使用 Trimmomatic v0.38 软件（Bolger et al., 2014）过滤原始 reads 数据（参数：HEADCROP:15 LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:50 AVGQUAL:20），再使用 Trinity v2.6.6 软件（Grabherr et al., 2011）拼接（min contig length=150bp），最后使用 TransDecoder v5.5.0（https://github.com/TransDecoder/TransDecoder/releases/tag/TransDecoder-v5.5.0 ）进行 CDS 和蛋白质序列预测（参考数据库为 Swissprot 和 Pfam-A）。将得到的这两个物种的基因集、从 onekp 数据库下载得到的 43 种被子植物的基因集和 1 个裸子植物（*Ginkgo biloba*）的基因集，使用 HaMStR v13.2.6 软件（Ebersberger et al., 2009）合并到利用基因组数据得到的同源基因家族中，最终只保留大于 50 个样品的基因家族。

1.4 系统发育进化树构建

我们采用两种方法串联法（concatenation）和溯祖法（coalescence），并分别使用 CDS 序列和氨基酸序列构建进化树。无论是 CDS 序列还是蛋白质序列，都使用 PRANK v.170427 软件（http://wasabiapp.org/software/prank/ ）进行多序列比对，使用 PHYUTILITY v2.2.6 软件（Smith & Dunn, 2008）修剪缺失率大于 70%的位点，其中 CDS 序列需去除长度小于 300 个碱基的序列，蛋白质序列需去除长度小于 100 个氨基酸的序列。

溯祖法，先对每个基因使用 RAxML v8.2.11 软件（默认参数）（Stamatakis, 2014）画树，

再使用 ASTRAL v5.5.9 软件（Zhang et al., 2017）处理所有基因树，得到共有树，参数设置“-t 1 --gene-only”以获得 bootstrap 值和基因支持率，枝长使用 iqtree v1.5.5 软件（Nguyen et al., 2015）获得。

串联法，先使用 PartitionFinder v2.1.1 软件（Lanfear et al., 2009）对串联序列进行分区和进化模型检测，从而设置较合理的分区和为每个分区选择合理的进化模型。对 CDS 序列检测下列的四个分区策略（表 2）:no partitioning, partitioning by each codon position (three partitions), partitioning by gene 和 partitioning by each codon position within each gene。对蛋白质序列检测下列两个分区策略：no partitioning 和 partitioning by gene。参数设置如下：branch lengths = linked；model_selection = aicc；search = user；models = GTR, GTR+G, GTR+I+G（CDS 序列）或者 models = LG+G, LG+I+G, WAG+G, WAG+I+G（蛋白质序列）。再使用 iqtree v1.5.5 软件画树（1000 ultrafast bootstrap replicates（Von Haeseler et al., 2013），-spp 设置最优分区策略），基因支持率使用 ASTRAL v5.5.9 软件（-t 1）获得。最后使用软件 Evolview v2（He et al., 2016）对获得的所有进化树进行美化。

表 2 串联法建树分区模型检测

Table2 AICc scores for each of the phylogenetic matrix partitioning strategies

数据 Matrix	数据大小 Number of data	分区策略 Partitioning strategy	分区数目 Number of partitions	对数似然值 Log-likelihood	赤池信息值 AICc
nt(≥50sample)	26 563 047	OnePart	1	-343 591 104.000 000	687 182 578.003 000
		CodonPart	3	-343 585 496.000 000	687 171 406.003 000
		GenePart	5 929	-342 778 283.430 175	685 687 673.619 000
		CodonGenePart	**3×5 929**	**-341 770 443.474 243**	**683 935 206.166 000**
nt(≥70sample)	16 540 374	OnePart	1	-222 849 712.000 000	445 699 794.004 000
		CodonPart	3	-222 846 000.000 000	445 692 414.005 000
		GenePart	3 384	-222 406 333.714 843	444 887 632.932 000
		CodonGenePart	**3×3 384**	**-221 758 883.619 628**	**443 742 935.528 000**
nt(≥80sample)	9 340 788	OnePart	1	-129 962 472.000 000	259 925 314.007 000
		CodonPart	3	-129 960 500.000 000	259 921 414.009 000
		GenePart	1 791	-129 739 122.166 992	259 518 079.097 000
		CodonGenePart	**3×1 791**	**-129 354 388.143 432**	**258 828 073.274 000**
nt(≥85sample)	4 069 848	OnePart	1	-57 607 360.000 000	115 215 090.017 000
		CodonPart	3	-57 606 964.000 000	115 214 342.021 000
		GenePart	742	-57 512 358.070 313	115 041 422.363 000
		CodonGenePart	**3×742**	**-57 329 703.801 392**	**114 709 026.252 000**
nt(≥89sample)	231 309	OnePart	1	-3 311 701.250 000	6 623 772.797 760
		CodonPart	3	-3 311 505.937 500	6 623 426.247 620
		GenePart	42	-3 304 863.765 625	6 611 003.043 870
		CodonGenePart	**3×42**	**-3 290 917.219 238**	**6 584 975.637 010**
AA(≥50sample)	3 332 638	OnePart	1	-135 739 947.205 826	271 508 915.612 567
		GenePart	**5 929**	**-135 248 986.841 308**	**270 526 876.482 000**
AA(≥70sample)	2 029 014	OnePart	1	-83 878 051.077 318	167 759 844.588 965
		GenePart	**3 384**	**-83 574 671.167 480**	**167 169 596.938 000**
AA(≥80sample)	1 165 765	OnePart	1	-47 214 500.000 000	94 429 354.054 100
		GenePart	**1 791**	**-47 043 728.832 520**	**94 097 113.497 200**

AA(≥85sample)	519 158	OnePart	1	-19 925 224.000 000	39 850 802.121 400
		GenePart	**742**	**-19 851 832.912 842**	**39 707 977.665 800**
AA(≥89sample)	30 148	OnePart	1	-979 643.625 000	1 959 681.828 340
		GenePart	**42**	**-973 566.588 867**	**1 948 059.215 120**

注：黑体为最优模型（即赤池信息值值最低）

Note: Bold is the best partition (AICc value is the lowest).

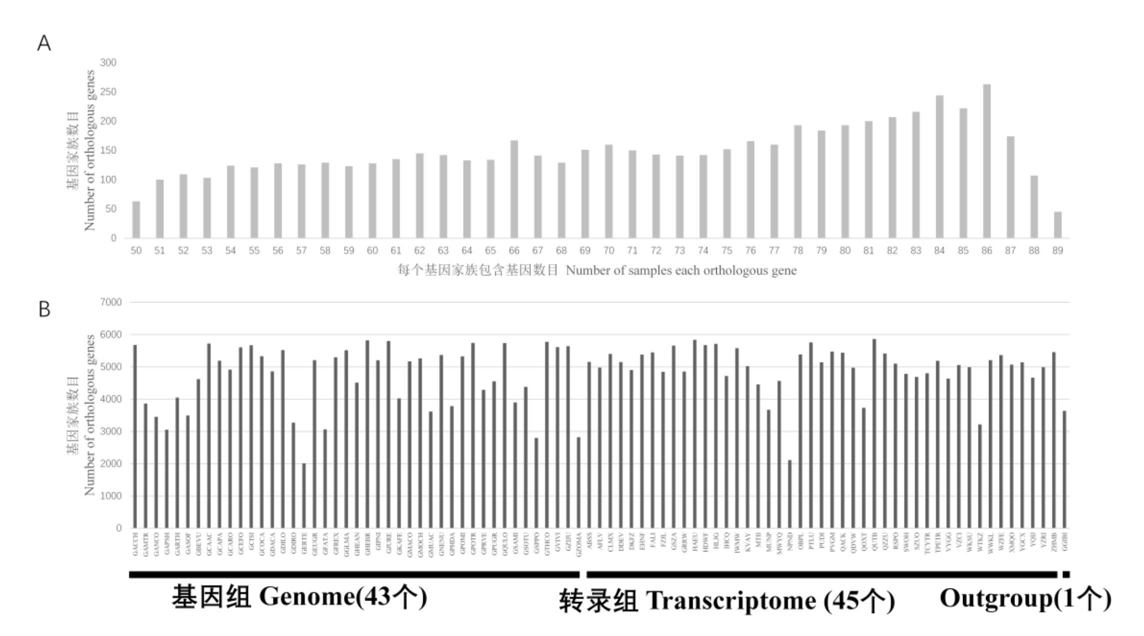
1.5 分化时间估计

我们使用 PAML v4.9 软件包（Yang, 2007）的 MCMCTREE 程序进行分化时间估计，输入拓扑结构为综合 20 棵进化树的最佳拓扑结构（即使用 742 个基因的 CDS 序列串联法获得的拓扑结构），输入序列为 742 个基因的 CDS 序列。我们先对每个基因都分别估计分化时间，再综合 742 个基因的分析结果（即每个节点取所有基因的平均值）获得最终的分化时间树。拓扑结构的枝长使用 JONES+gamma 碱基替换模型获得；rgene gamma 设定为 G(1, 4.5)；sigma2 gamma 设定为 G(1, 4.5)；clock 设定为 3；Markov chain Monte Carlo(MCMC)设定为 burnin = 50 000, sampfreq = 100, nsample = 10 000。对每个基因，都是分别运行两次独立的 MCMC（即不同的 random seeds），使用 Tracer v1.7 软件 (https://github.com/beast-dev/tracer/releases/tag/v1.7.1)观察运行结果是否稳定和收敛，所有节点及参数的 effective sample size 是否大于 200。九个化石校准设定为：银杏分化时间为 290-310 百万年前（Gao et al., 1989），单子叶植物和真双子叶植物分化时间为 130~200 百万年前（Kumar et al., 2017），真双子叶植物共同祖先（即最早的双子叶植物化石记录）为 125 百万年前（Herendeen, 1995; Zeng et al., 2014），山龙眼目（Proteales）的共同祖先为 108.8 百万年前（Crane et al., 1996），葡萄目（Vitales）与其余蔷薇类植物间分化时间为 105-115 百万年前（Fawcett et al., 2009; Kumar et al., 2017），*A. thaliana* 与 *P. trichocarpa* 间分化时间为 97~109 百万年前（Kumar et al., 2017），豆目（Fabales）与壳斗目（Fagales）间分化时间为 93.5 百万年前（Friis et al., 1996），山茱萸目（Cornales）共同祖先为 85.8 百万年前（Takahashi et al., 2002），唇形目（Lamiales）共同祖先为 44.3 百万年前（Call et al., 1992）。

2 结果与分析

2.1 直系同源基因鉴定

我们对 44 个植物基因组基因集和 45 个已拼接转录组 CDS 序列进行同源基因聚类，并使用 Yang & Smith(2014)报道的方法，去除所有旁系同源基因，最终获得大于 50 个样品的 one-to-one 基因家族（即每个样品最多一条序列）共 5_993 个（图 3:A），各种植物的基因覆盖率从 33.57%到 97.85%，平均为 80.40%（图 3:B）。



注：**A**. 每个同源基因家族含有的基因数目；**B**. 每个样品含有的同源基因家族数目。

Note：**A**. Number of samples for each orthologous gene family; B. Number of orthologous genes for each sample.

图 3 5 993 个聚类的同源基因家族

Fig.3 Results of 5 993 inferred orthologous gene groups

2.2 系统发育进化树构建

我们采用串联和溯祖法共构建了 20 棵进化树，并比较它们之间的不同（图 4），以评估树的稳定性。CDS 序列和蛋白质序列，都分别使用五个数据集，总共构建 20 棵树（5 棵 CDS 串联法树，5 棵 CDS 溯祖法树，5 棵 AA 串联法树和 5 棵 AA 溯祖法树）。这 5 个数据集分别包含 5 928 个 orthologs（≥50 samples）、3 384 个 orthologs（≥70 samples）、1 791 个 orthologs（≥80 samples）、742 个 orthologs（≥85 samples）及 42 个 orthologs（≥89 samples）。

这 20 棵进化树主要是为了进一步确定图 1 中五类被子植物间演化关系和真双子叶植物内部各目间系统发育关系。这些进化树中的大多数，是与使用 742 个基因 CDS 序列（共 4 069 848 位点）串联方法建立的进化树高度一致的（图 5）（使用 3 384 个基因 AA 序列建立的进化树，和使用 1 791 个基因 AA 序列建立的进化树，也是相同的最佳拓扑结构）。

2.2.1 木兰类植物、单子叶植物及双子叶植物间演化关系

无论核酸序列还是蛋白质序列，使用串联法和溯祖法建立的进化树基本都支持拓扑结构（(真双子叶植物，单子叶植物)，木兰类植物）（图 4）。

2.2.2 金粟兰科与金鱼藻科

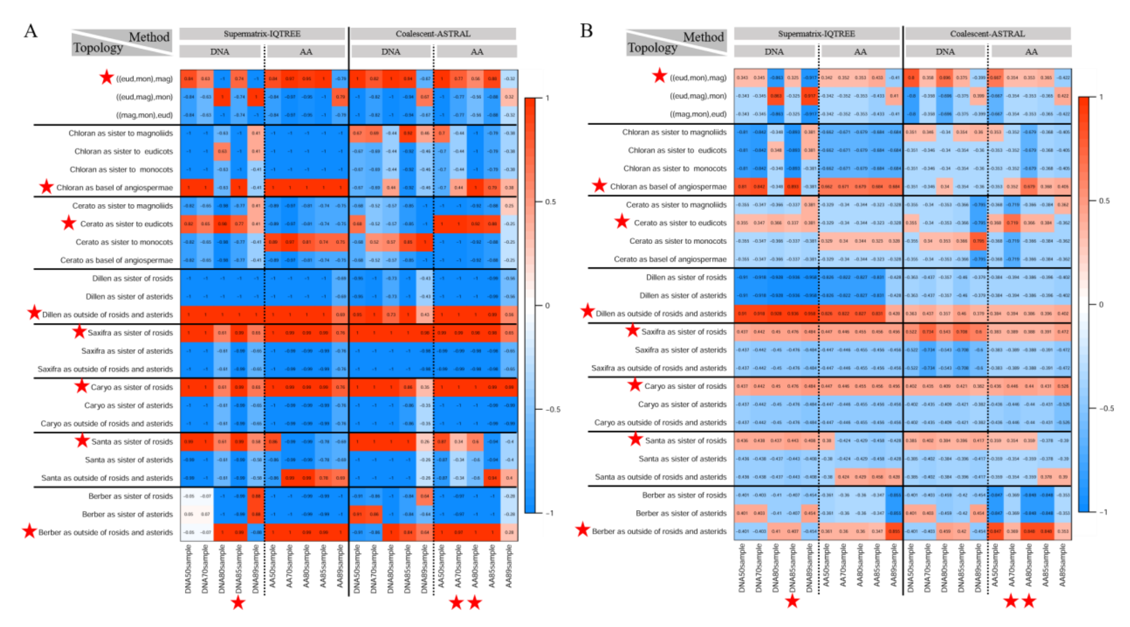
我们的研究表示，金鱼藻科是真双子叶植物的姊妹群，这与前人的研究结果一致(图 4)。但金粟兰科是所有被子植物（除 ANITA 外）的基底旁系群，这与 APG IV 认为的“金粟兰科是木兰类植物的姊妹群”是不同的。

2.2.3 双子叶植物内部各目的系统发育关系

我们的研究认为，五桠果科是蔷薇类植物和菊类植物共同的姊妹群，虎耳草目是蔷薇类植物的姊妹群，这都与 APG IV 一致（图 4）。

APG IV 认为“檀香目和石竹目是菊类植物的姊妹群”，而我们的研究否定了这一结论：20 棵进化树中，所有结果都支持“石竹目是蔷薇类植物的姊妹群”；大部分支持“檀香目是蔷薇类植物的姊妹群”，这与 Zeng et al.(2017)等的研究结果一致；少部分支持“檀香目是蔷薇类植物和菊类植物共同的姊妹群”（图 4）。

APG IV 认为“智利藤目是菊类植物的姊妹群”，而我们的研究只有少部分支持这一结论。使用蛋白质序列建立的进化树，无论串联还是溯祖法，都支持“智利藤目是蔷薇类植物和菊类植物共同的姊妹群”。使用核酸序列建立的进化树，随着基因数目的增多，逐渐转变为支持“智利藤目是菊类植物的姊妹群”，与 APG IV 一致（图 4）。

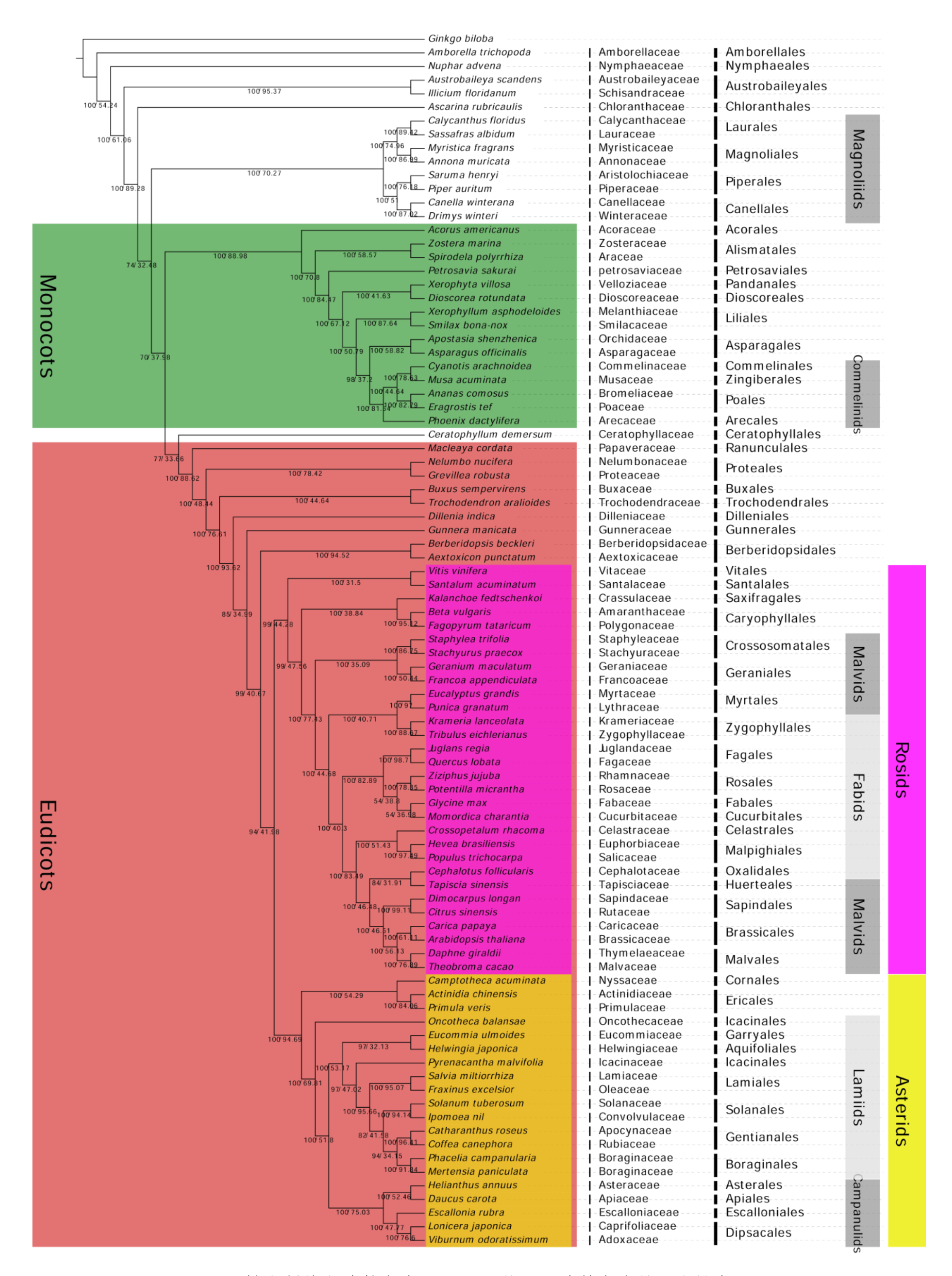


注：**A**. Bootstrap 值；**B**. 基因支持率。(红色表示支持，蓝色表示拒绝，红色星星表示支持率最高的拓扑结构)

Note：**A**. Bootstrap values; **B**. Gene trees support values. (Red represents support, blue represents rejection, the topology labeled with a red star is the most supported.)

图 4 采用各种数据集并使用串联和并联方法建立的 20 棵进化树，对各种有争议拓扑结构的支持率统计

Fig.4 Statistics of support values for relationships among the clades which are controversial in previous studies by using different methods and gene numbers



注：枝上斜线左边数字为 bootstrap 值，右边数字为基因支持率。

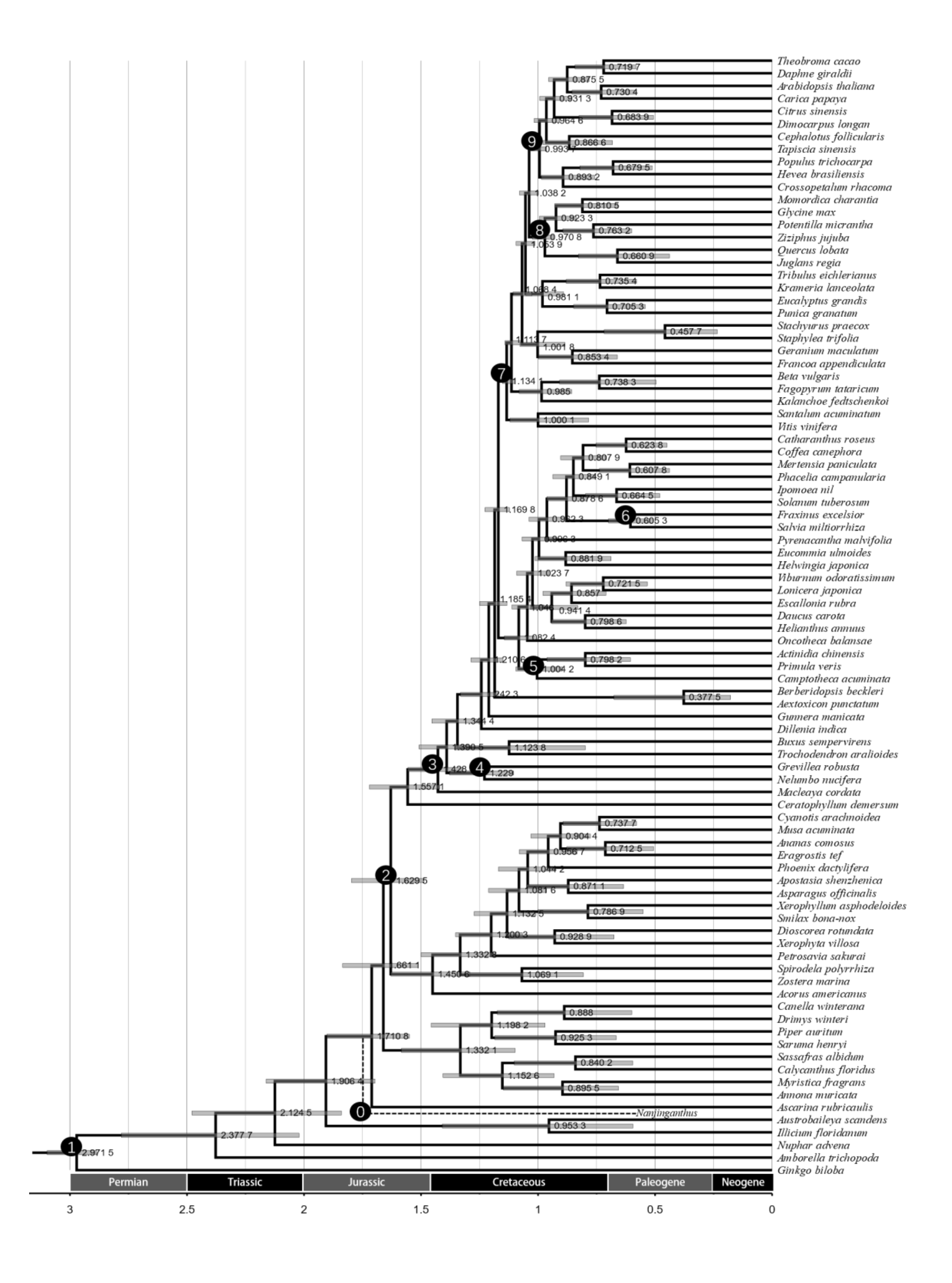
Note：The left number at the notes is bootstrap value. The number on the right is gene trees support ratio.

图 5 使用 742 个基因 CDS 序列串联方法构建的系统发育进化树

Fig.5 Concatenation-based angiosperm phylogenetic tree based on CDS sequences of 742 orthologs

2.3 分化时间估计

基于 742 个基因 CDS 序列串联方法建立的进化树，我们估计了被子植物的分化时间(图 6)。我们认为被子植物的起源时间为 237.78 百万年前（95%置信区间为 202.6~278.08），与主流观点认为的 225~240 百万年前一致 (Magallon, 2010; Smith et al., 2010; Zeng et al., 2014)。木兰类植物与单子叶植物和真双子叶植物的分化时间约为 166.11 百万年前；五桠果科与蔷薇类和菊类植物的分化时间约为 124.23 百万年前；蔷薇类植物与菊类植物的分化时间约为 116.98 百万年前；唇形类植物（Lmiids）与桔梗类植物（Campanulids）的分化时间约为 102.37 百万年前。



注：灰色条纹为分化时间的 95%置信区间,九个化石校准时间为：（1）银杏分化时间为 290-310 百万年前；（2）单子叶植物和真双子叶植物分化时间为 130~200 百万年前；（3）真双子叶植物共同祖先（即最早的双子叶植物化石记录）为 125 百万年前；（4）山龙眼目（Proteales）的共同祖先为 108.8 百万年前；（5）山茱萸目（Cornales）共同祖先为 85.8 百万年前；（6）唇形目（Lamiales）共同祖先为 44.3 百万年前；（7）葡萄目（Vitales）与其余蔷薇类植物间分化时间为 105-115 百万年前；（8）豆目（Fabales）与壳斗目（Fagales）间分化时间为 93.5 百万年前；（9）*Arabidopsis thaliana* 与 *Populus trichocarpa* 间分化时间为 97~109 百万年前；（0）"南京花"的可能系统演化位置（175 百万年前）。

Note：Grey bars are 95% confidence intervals, nine fossil calibration points are as follows: (1)The divergence time of *Ginkgo biloba* is 290-310 million years ago. (2) The divergence time of eudicots and monocots is 130-200 million years ago. (3) The divergence time of eudicots is 125 million years ago. (4) The divergence time of Proteales is 108.8 million years ago. (5) The divergence time of Cornales is 85.8 million years ago. (6) The divergence time of Lamiales is 44.3 million years ago. (7) The divergence time of Vitales from Rosids is 105-115 million years ago. (8) The divergence time of Fabales and Fagales is 93.5 million years ago. (9) The divergence time of *Arabidopsis thaliana* and *Populus trichocarpa* is 97-109 million years ago.

图 6 基于 742 个基因 CDS 序列对被子植物分化时间的估计结果

Fig.6 Chronogram presenting estimated divergence times by MCMCTREE using CDS sequences of 742 genes

3 讨论与结论

长期以来，被子植物的系统发育关系重建，都是使用质体基因、线粒体基因或少数保守的单拷贝核基因。Yang & Smith(2014)报道了一种基于系统进化树的同源基因聚类及去旁系同源基因的方法，我们使用此种方法对收集的 88 种植物核基因集进行聚类，共获得了多达 5 993 个 one-to-one 基因家族，并从这个数据集里面截取各种大小的数据进行进化树重建，以测定进化树的稳定性。

获得比以前更多的核基因家族后，制约系统演化关系构建的另一个因素就是大量的计算资源和计算时间。构建系统进化树时，一般需要设置 bootstrap 值（100~1 000）迭代，此步骤非常耗费计算时间。Nguyen et al.(2015)发表的软件 iqtree，采用 ultrafast bootstrap approximation(UFBoot)方法获得 bootstrap 值(Von Haeseler et al., 2013)，比 RAxML 软件的传统方法，计算速度快 10~40 倍，并且获得的 bootstrap 值更精确。

我们使用多达 5 993 个 one-to-one 基因家族构建的进化树，与 APG IV 报道的主要差异为檀香目和石竹目在系统发育树中的位置，本研究认为"檀香目和石竹目是蔷薇类植物的姊妹群"，而 APG IV 认为"檀香目和石竹目是菊类植物的姊妹群"。可能原因有以下两个：一是基因数目的增多；二是本研究所选 88 个植物只有一半使用的基因组序列，另一半为转录组序列，而转录组序列一般存在大量的基因缺失（即未表达基因较多）。

总的来说，本研究不仅进一步确定了被子植物各目间系统发育关系，而且为"使用更多的基因和计算速度更快的方法构建进化树"探讨了一种可行性策略：即使用 Yang & Smith(2014)报道的同源基因聚类及去旁系同源基因方法，获得大量的 one-to-one 基因家族，再使用 IQ-TREE（串联法）和 ASTRAL（溯祖法）软件，能快速精确的计算出进化树。随着更多植物基因组的测序和基因聚类及系统发育关系构建方法的进一步优化，被子植物系统发育关系将越来越精确，例如进一步准确确定檀香目和石竹目在被子植物中与其他进化分支之间的关系。

参考文献：

BOLGER AM, LOHSE M, USADEL B, 2014. Trimmomatic: A flexible trimmer for Illumina sequence data[J]. Bioinformatics, 30(15): 2114-2120.

CALL VB, DILCHER DL, 1992. Investigations of angiosperms from the Eocene of southeastern North America: Samaras of *Fraxinus wilcoxiana* Berry[J]. Rev Palaeobot Palynol, 74: 249-266.

CHAW SM, LIU YC, WU YW, et al., 2019. Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution[J]. Nat Plants, 5(1): 63-73.

CHEN JH, HAO ZD, GUANG XM, et al., 2019. Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation[J]. Nat Plants, 5(1): 18-25.

CRANE PR, HERENDEEN PS, 1996. Cretaceous floras containing angiosperm flowers and fruits from eastern North America[J]. Rev Palaeobot Palynol, 90: 319-337.

EBERSBERGER I, STRAUSS S, VON HAESELER A, 2009. HaMStR: Profile hidden markov model based search for orthologs in ESTs[J]. Bmc Evol Biol, 9(1): 157-157.

ENDRESS PK, DOYLE JA, 2009. Reconstructing the ancestral angiosperm flower and its initial specializations[J]. Amer J Bot, 96(1): 22-66.

FAWCETT JA, MAERE S, VAN DE PEER Y, 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event[J]. Proc Nat Acad Sci USA, 106(14): 5737-5742.

FRIIS EM, PEDERSEN KR, SCHÖNENBERGER J, 2006. Normapolles plants: A prominent componentof the Cretaceous rosid diversification[J]. Plant Syst Evol, 260: 107-140.

FU Q, DIEZ JB, POLE M, et al., 2018. An unexpected noncarpellate epigynous flower from the Jurassic of China[J]. Elife, 7: e38827.

GAO Z, BARRY AT, 1989. A review of fossil cycad megasporophylls, with new evidence of *Crossozamia pomel* and its associated leaves from the lower permian of Taiyuan, China[J]. REV Palaeobot Palynol, 60(3-4): 205-223.

GRABHERR MG, HAAS BJ, YASSOUR M, et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. Nat Biotechnol, 29(7): 644-652.

HE ZL, ZHANG HK, GAO SH, et al., 2016. Evolview v2: An online visualization and management tool for customized and annotated phylogenetic trees[J]. Nucl Acid Res, 44(W1): 236-241.

HERENDEEN PS, 1995. The enigma of angiosperm origins[J]. Earth-Sci Rev, 39(1): 253-254.

KATOH K, STANDLEY DM, 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability[J]. Mol Biol Evol, 30(4): 772-780.

KUMAR S, STECHER G, SULESKI M, et al., 2017. TimeTree: A Resource for 598 Timelines, Timetrees, and Divergence Times[J]. Mol Biol Evol, 34: 1812-1819.

LANFEAR R, FRANDSEN PB, WRIGHT AM, et al., 2016. PartitionFinder 2: New methods for selecting partitioned models of evolution formolecular and morphological phylogenetic analyses[J]. Mol Biol Evol, 34(3): 772-773.

LU LM, MAO LF, YANG T, et al., 2018. Evolutionary history of the angiosperm flora of China[J]. Nature, 554(1): 234-238.

LI HT, YI TS, GAO LM, et al., 2019. Origin of angiosperms and the puzzle of the Jurassic gap. Nat Plants, 5(1): 461-470.

MAGALLON S, 2010. Using fossils to break long branches in molecular dating: A comparison of

relaxed clocks applied to the origin of angiosperms[J]. Syst Biol, 59(4): 384-399.
MOORE MJ, HASSAN N, GITZENDANNER MA, et al., 2011. Phylogenetic Analysis of the Plastid Inverted Repeat for 244 Species: Insights into Deeper-Level Angiosperm Relationships from a Long, Slowly Evolving Sequence Region[J]. Int J Plant Sci, 172(4): 541-558.
MOORE MJ, SOLTIS PS, BELL CD, et al., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots[J]. Proc Nat Acad Sci USA, 107(10): 4623-4628.
NGUYEN LT, SCHMIDT HA, VON HAESELER A, et al., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. Mol Biol Evol, 32(1): 268-274.
QIU YL, LI LB, WANG B, et al., 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes[J]. JSE, 48(6): 391-425.
RUHFEL BR, GITZENDANNER MA, SOLTIS PS, et al., 2014. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes[J]. Bmc Evol Biol, 14(1): 23.
SMITH SA, BEAULIEU JM, DONOGHUE MJ, 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants[J]. Proc Nat Acad Sci USA, 107(13): 5897-5902.
SMITH SA, DUNN CW, 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data[J]. Bioinformatics, 24(5): 715-716.
SOLTIS DE, SMITH SA, CELLINESE N, et al., 2011. Angiosperm phylogeny: 17 genes, 640 taxa[J]. Amer J Bot, 98(4): 704-730.
STAMATAKIS A, 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies[J]. Bioinformatics, 30(9): 1312-1313.
TAKAHASHI M, CRANE PR, MANCHESTER SR, 2002. Hironoia fusiformis gen. et sp. nov., A cornalean fruit from the Kamikitaba locality (Upper Cretaceous, Lower Coniacian) in northeastern Japan[J]. J Plant Res, 115: 463-473.
THE ANGIOSPERM PHYLOGENY GROUP, 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV[J]. Bot J Linn Soc, 181(1): 1-20.
VAN DS, 2000. Graph Clustering by Flow Simulation[M]. University of Utrecht.
VON HAESELER A, MINH BQ, NGUYEN MAT, 2013. Ultrafast Approximation for Phylogenetic Bootstrap[J]. Mol Biol Evol, 30(5): 1188-1195.
WICKETT NJ, MIRARAB S, NGUYEN N, et al., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants[J]. Proc Nat Acad Sci USA, 111(45): 4859-4868.
WORBERG A, QUANDT D, BARNISKE AM, et al., 2007. Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving DNA[J]. ORG DIVERS EVOL, 7(1): 55-77.
Yang Z, 2007. PAML 4: Phylogenetic analysis by maximum likelihood[J]. Mol Biol Evol, 24:1586–1591.
YANG Y, MOORE MJ, BROCKINGTON SF, et al., 2015. Dissecting Molecular Evolution in the Highly Diverse Plant Clade Caryophyllales Using Transcriptome Sequencing[J]. Mol Biol EVOL, 32(8): 2001-2014.
YANG Y, SMITH SA, 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics[J]. Mol Biol Evol, 31(11): 3081-3092.

ZENG LP, ZHANG N, ZHANG Q, et al., 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets[J]. New Phytol, 214(3): 1338-1354.

ZENG LP, ZHANG Q, SUN RR, et al., 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times[J]. Nat Comm, 5(1): 4956.

ZHANG C, SAYYARI E, MIRARAB S, 2017. ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches[J]. RECOMB-CG, Springer, Cham: 53-75.

ZHANG N, ZENG LP, SHAN HY, et al., 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms[J]. New Phytol, 195(4): 923-937.